**Abstract:** In the 1970s, Dr. Bill Gray proposed that tropical cyclone (TC) (strong TCs in the Atlantic are known as hurricanes) formation was strongly influenced by several large scale environmental factors (LSEFs): high sea surface temperature, low vertical wind shear, upward moving air, positive vorticity, and high humidity. Since then, observations of hundreds of storms have reinforced Gray’s findings, but with the caveat that the LSEFs are necessary but not sufficient for TC formation. Since its inception in 2007, Statistical Solutions LLC, in collaboration with the Naval Postgraduate School, has built upon these results to:

1. Test LSEFs for statistical significance
2. Develop a statistical model that relates the LSEFs to TC formation
3. Force the model with dynamical weather model forecasts of the LSEFs at leads ranging from 1 day to 90+ days to create probabilistic forecasts of TC formation
4. Evaluate those forecasts for skill

In this paper, we discuss the process of significance testing of the predictor LSEFs, statistical model development, selection of optimal dynamical model outputs, and visualization of the outputs. We also show examples of TC formation forecasts at various leads, and close with a discussion of the challenges of forecast verification.

**Tracks:** Data Mining, Probability and Applied Statistics, Local Companies in OR, Military OR, Applications, Modeling, Visualization, Tropical cyclone formation, Typhoons, Hurricanes, Dynamical-Statistical Forecast

1. **Introduction**

There is an obvious need for improved understanding of tropical cyclone (TC) activity, in particular, how that activity is influenced by the surrounding environment. While TC activity is comprised of three components --- formation, track and intensity, we focus on formation because improving our understanding of formation is a logical first step in improving our understanding of TC activity overall.

TC formation is widely recognized to be more likely when conditions are favorable in several large-scale environmental factors (LSEFs). These favorable LSEF conditions are considered to be necessary but not sufficient for TC formation [Gray, 1975; Frank, 1987; McBride, 1995]. Though

---

1 Statistical Solutions LLC, 3746 Mesquite Dr, Beavercreek, OH 45440 (david.statistical.solutions@gmail.com, http://www.statisticalsolutionsllc.com/tropical-cyclone-forecasts---2015.html)
2 Department of Meteorology, Naval Postgraduate School, 254 Root Hall, 589 Dyer Rd, Monterey, CA, 93943-5114 (murphree@nps.edu)
different studies have proposed slightly different sets of favorable LSEFs, most studies agree on and include some form of the following:

- Sea surface temperatures (SSTs) above 26°C
- Weak shear between upper and lower level horizontal winds
- Positive absolute vorticity at low levels
- Mean upward motion
- High mid-level humidity

While there is general consensus that favorable LSEFs influence TC formation, more advanced analyses of the LSEF-TC formation relationships are needed. The LSEF data used by early investigators of tropical cyclogenesis consisted of a combination of climatological, weather service summary, and in situ observational data [Gray, 1968]. Since those early studies, improvements in the types, availability, quality, and resolution of environmental data, as well as data sets populated with that much more data than that used in prior studies allows us to explore the LSEF-TC formation relationship in depth. This exploration ultimately enabled the development of a statistical model that yields the probability of TC formation given what the values of the LSEFs are at a specified time and location.

Figure 1 shows the sea surface temperature (SST) for 01 January 1979 from the Climate Forecast System Reanalysis (CFSR) generated by the National Centers for Environmental Prediction (NCEP) [Saha et al, 2010]. This figure represents the type of long term climate system data presently available for analyzing relationships between climate system variables and for developing forecasting systems for predicting future states of the climate system. Similar data is available for many atmospheric, oceanic, and land variables for multiple decades at hourly temporal resolution, one-half degree horizontal resolution, and many levels within the ocean and atmosphere. The LSEF variables are available either directly from the CFSR data set or may be derived directly from other CFSR variables (such as shear and vorticity).

Figure 1. Sea surface temperature (SST; degrees C) for 01 January 1979 from the CFSR data set. CFSR data is available from 1979 through the present, with global coverage.
We obtained TC formation data (time and location) from the Joint Typhoon Warning Center (JTWC). Annually, JTWC publishes their best track analysis, which consists of reanalyzed data on the time, location, strength, and other information for each TC, by basin (e.g., the western North Pacific (WNP), the northern Indian Ocean, etc.). For each TC, we used the first time and position listed in the best track analysis as the time and location of TC formation. Figure 2 shows the formation locations of all June through November WNP TC formations for the year 2010.

Figure 2. The locations of TC formations (shown by pink dots) in the western North Pacific (WNP) during June-November 2010 based on the first data entry for each TC in the JTWC best track analysis for 2010. The best track analyses extend from 1945 to the present.

2. Data and Methods

The basin used for our study was the western North Pacific (WNP). We chose this region to maximize the number of formations in the data set (the WNP accounts for nearly 30% of the global annual total of TCs [Chan, 2004] (see Figure 3, [Nilfanion, 2006]). We define the WNP as the region bounded by: 0-30° North and 100-180° East, consistent with areas defined as the WNP by other researchers [Chan and Liu, 2004]. We chose 1979-2009 as our initial study period to maximize the amount and consistency of data available (at the time we developed the TC formation statistical model, 2009 was the most recent year for which JTWC best track data was available). The data we used for statistical model building was limited to that from June-November, which are the months when TC activity tends to be greatest in the WNP [Frank, 1987].
Figure 3. The tracks and intensities of TCs during 1985-2005. The colored lines indicate the tracks of individual TCs, with warmer colors indicating higher intensities (greater wind speeds). More TCs occur in the western North Pacific (WNP) than in any other basin.

Our tool of choice for the development of the statistical model was logistic regression. The LSEFs used in the regression model as predictors of TC formation were slightly different from those described in section 1 to account for data quality and availability issues and to facilitate analysis of LSEF-TC formation relationships. The LSEF predictors we used were SST, shear, relative humidity, upper level divergence (as a proxy for vertical motion), planetary vorticity and relative vorticity (the components of absolute vorticity). The exact nature of the individual LSEF-TC formation relationships are not well understood, so we included in the initial model second order polynomial terms of all predictors, and second order interaction terms. Ultimately, using backward stepwise logistic regression [Collett, 1991], we determined that the probability of TC formation is a function of (at over a 0.999 confidence level) SST and SST squared, shear and shear squared, divergence, relative vorticity, and three secondary interaction terms. Planetary vorticity by itself is not significant, but it is part of one of the significant interaction terms and thus was retained as a predictor. It was also determined in the initial model building process that multicollinearity was an issue (it turns out that warm SSTs, upward air motion, and positive upper level divergence tend to be strongly associated with high relative humidity. Therefore, we dropped the humidity term and re-ran the logistic regression to create a model that was physically plausible. Cross-validation was used as a check against over-fitting. In addition, two years of independent LSEF and TC formation data since the last update of the statistical model in winter 2014 have produced results that indicate that the model is valid.

One example of the statistical model’s skill is provided by Figure 4 which shows the statistical model calculated long term mean (LTM) probability of TC formation in the WNP for July-October based on best track and LTM LSEF data for 1979-2009 (colored contours), along with the formation locations for all the TCs that occurred during July-October of 1979-2009 (blue dots) [Johnson, 2011]. The good agreement between the TC formation distribution and the statistical model contours provides some evidence that the model has skill. The figure does not show TC formations east of 180E as they are considered central North Pacific storms.
Figure 4. Long term mean probabilities of TC formation generated by the statistical model (colored contours) and actual TC formations (blue dots) for July-October 1979-2009.

The focus of our forecasting, though, has been on TC formation forecasts at extended leads, beyond the 5-10 day skill limit of dynamical models, and with the temporal and spatial resolution missing from seasonal forecasts (those forecasts that merely state an expected number of TC (or hurricane) formations for the year with little prediction of when or where those formations would occur).

To force the statistical model, we use extended lead forecasts of the LSEFs from the NCEP Climate Forecast System version 2 (CFSv2) \cite{Saha et al, 2014}, which uses the same dynamical model as that used to produce CFSR and has the same spatial and temporal resolution as the CFSR data used to build the statistical model. We felt that if the CFSv2 forecasts of the LSEFs are at least reasonably skilled, then skilled extended lead forecasts of TC formation probabilities could be made by forcing the statistical model with the CFSv2 forecasts.

A motivation for trying this approach is that CFSv2 and other extended range dynamical models tend to be good at forecasting large-scale features, such as SSTs or winds, but are less skilled at forecasting synoptic scale events (such as tropical cyclogenesis) beyond 5-10 days. Conversely, statistical models, though not used as much as dynamical models for forecasting large-scale features, are well suited to forecasting probabilities of synoptic scale events. Thus, we developed the Statistical Solutions LLC-Naval Postgraduate School TC Formation Forecasting System (hereafter referred to as TCFS) to capitalize on the strengths of both dynamical and statistical models.

The CFSv2 forecast system is an ensemble forecast system, meaning that it produces multiple forecasts, or an ensemble of forecasts, for a single future time frame \cite{Saha et al, 2014}. The CFSv2 forecasts we use are issued four times daily at 00, 06, 12, and 18z. A major motivation for ensemble forecasting is to account for the uncertainty in the initial conditions for the forecast model (i.e., the observed conditions used at the start time for the model) as well as that imposed by the model itself. The forecasts for the different ensemble members can be very similar or very different. The amount of difference depends to a large extent on the sensitivity of the model to the initial conditions, which itself is determined in large part by the degree of nonlinearity in the modeled atmosphere. The amount of difference, and the range in outcomes from the forecasts, tends to increase as the lead time increases --- mainly because the effects of sensitivity and nonlinearity tend to increase with time. So at very short leads (e.g., 1 day), the effects of changing initial conditions, uncertainty in those initial conditions and model inaccuracies tend to be small, and the ensemble member forecasts tend to be very similar. 
longer leads, the spread in each ensemble member forecast, and thus the uncertainty in the forecasts, tends to increase. To overcome this issue, more ensembling is required.

The use of more ensemble members at longer lead times to increase the forecast skill is analogous to the use of Monte Carlo simulation. That is, a larger number of ensemble members tends to produce a more representative forecast distribution (as representative as forecasts can be) of the actual LSEF values at the valid time of the forecast. We found that the longer the forecast lead-time, the greater the amount of ensemble members needed.

Figure 5 shows a typical 1-day lead forecast, plus the location of a TC that formed during the forecast valid period (taken from the first position indicated on the corresponding JTWC TC formation alert). The 1-day lead forecasts are the TCFS forecasts with the highest skill (as discussed further in section 3). While 1-day lead forecasts do not sound terribly helpful, tropical cyclogenesis is a prolonged process, and formation is often not recognized until one or more days after formation occurs. So a forecast such as the one in Figure 5 may be issued as much as 1-3 days prior to a TC formation alert being issued, with the date and time of TC formation subsequently being identified as occurring 1-3 days prior to the alert being issued. Thus, the 1-day lead forecasts are useful for TC situational awareness in alerting forecasters and others where to focus their attention and where they may be able to devote less attention.

Similar to the 1-day lead forecasts are the 4-day lead forecasts (not shown). Like the 1-day forecasts, the dynamical model output itself is both skilled and little impacted by the changing initial conditions or model inaccuracies (though more so than the 1-day lead forecasts). Thus, light ensembling (8 members) is required to have coherent, and skilled forecasts.

The next (in lead time) TCFS products are the Weekly Outlooks. Figure 6 shows a Week 1 Weekly Outlook (Week 1 products are issued Mondays and valid for Wed-Tues), Week 2 Outlooks, valid
for the week following the Week 1 Outlook (not shown) are also produced. The issued date and the valid period line up with CPC’s production schedule for their weekly Global Tropical Benefits/Hazards Forecasts (GTHB) as we provide WNP TC formation technical input for those forecasts. In addition to the convenience for CPC to receive a forecast with a valid period that matches their own products, the longer valid period is necessary because as lead times grow, forecast errors and uncertainties in the LSEFs become more pronounced, making daily resolution forecasts impractical. Additionally, some amount of ensembling must be used to provide skilled, representative LSEF input. Therefore, after some experimentation to find the optimal amount of ensembling to maximize skill without washing out important synoptic scale climate events, we settled on Weekly Outlooks using 28 ensemble members per LSEF as the most practical and skilled solution to creating forecasts at intermediate leads. However, with that ensembling there are two things that occur: the amount of area that is shown as favorable for TC formation (i.e. that portion of the plot that is colored or contoured) becomes much larger than would be observed for a 1 or 4-day lead forecast due to the many varied inputs, and the sharpness or resolution of the forecast contours gets flatter in comparison to the daily forecasts. This is because of the many varied inputs and averaging that occurs, making determination of where formation may occur more challenging.

Figure 6. Week 1 Weekly Outlook of TC formation probability, with formation location of 8W (Kujira) identified with pink dot. Note the consistency between the highest formation probabilities seen in Figure 4 and those shown here, even though the LSEF forecasts to create the 1 and 4-day lead forecasts are different than those used to create the forecast shown here.

Figure 7 shows one of the TCFS monthly outlooks. It is valid for June of 2015, and was issued in late March, indicating the forecast was issued at a 2-month lead. The 3 TCs that formed in the WNP in June are shown. In looking at Figure 7, two points are apparent: first, the contoured area shown in Figure 7 in comparison to the climatology shown in Figure 3 is a substantial improvement in forecasting where formation may or may not occur. Second, there is good consistency between the Monthly Outlook of Figure 7, the Weekly Outlook of Figure 6, and the 1-day lead forecast of Figure 5. While the consistency is hardly proof of the skill of the TCFS, substantial inconsistency would cast doubt on the skill of the
system. Additionally, while forecasters tend to focus on skill, anything that can discriminate where formation may occur is important to planners. Though the formation favorable region of Figure 7 is sizeable, an example of its value is that it indicates (with the highly favorable region in the South China Sea (SCS)) that Vietnam and China are at substantial risk despite the fact that in a typical TC season, many months may pass without an actual SCS TC forming. This is particularly valuable information because June marks the beginning of the TC season, TC formations are relatively few [Mundhenk, 2009], and to skillfully issue a strong forecast for formation in the SCS for June is in stark contrast with what might be expected from climatology.

Because of the extended lead times (beyond a couple of weeks), monthly outlooks require the greatest amount of ensembling. Like the weekly outlooks, the ensembling tends to flatten the contours as compared to the short lead forecasts. However, an advantage of the added ensembling is that the amount of contoured area or area considered favorable for formation is smaller than that of the weekly forecasts because the contribution of individual ensemble member forecasts that are on the periphery is proportionally smaller than those with less ensembling, and thus is more likely to not meet the minimum threshold necessary for formation.

Figure 7. 2-Month Lead Monthly Outlook of TC formation probability, with formation location of 8W (Kujira), 9W (Chan-Hom), and 10W (Linfa) identified with pink dots. Of note: 10W, which only appears to be touching the contoured area is still counted as a hit. A neighborhood scoring system is used to account for inaccuracies in preliminary formation location, errors imposed by the grid, etc.

3. Verification of TCFS Performance

In the climatology world, a model is untrusted without verification. Figure 8 shows a variety of verification statistics for our Current Month Outlooks for 2013. The three statistics of most interest are the Probability of Detect (POD) – a measure of if a TC forms, how likely it is that it forms in a forecasted region; False Alarm Rate – a measure of TCFS forecasted formations but none occur, and Percent of Contoured Area (PCA). PCA is a verification statistic created by the authors of this paper. It is possible to improve POD by lowering the threshold of what is considered favorable for formation. But by doing
so, an increasingly larger contoured region results, reducing the discrimination (or the separation from climatology) that makes the forecasts useful. PCA then, is the ratio of the amount of contoured area in a forecast divided by the marine portion of the WNP that is considered possible for TC formation by climatology.

One note about the low value of FAR: FAR is zero because scores are calculated on a monthly basis. While the Monthly Outlook always indicated regions where conditions were favorable for formation, there was always at least one TC that formed during the valid period.

Figure 8. TCFS forecast skill metrics for our Current Month Monthly Outlooks, 2013. A skilled forecast would have positive a Heidke Skill Score (HSS) and Equitable Threat Score (ETS), high Percent Correct (PC) and POD, and a low PCA. These scores indicate that the TCFS at this lead is skilled.

Figure 8, (which shows typical verification statistics for any of our forecast years), and Figure 9, which is a bar chart of POD and PCA for all of the TCFS products for 2013 demonstrate that the TCFS has skill. It should also be noted that all forecasts and evaluations of forecasts are fully automated and objective (many TC forecasting systems require a subjective man-in-the-loop for forecasting and/or verification meaning the system is subject to the skill and objectivity of the man).
Figure 9. TCFS forecast skill metrics for all forecast products, 2013. These scores indicate that the TCFS is skilled at all leads.

5. Summary and Way Forward

Using over 30 years of climate (CFSR) and TC Best Track data, logistic regression was used to create a model that relates high-resolution LSEF values to the probability of TC formation. The necessary but not sufficient LSEFs first proposed by Gray [1968, 1975] were found to be statistically significant at a confidence level of over 0.999, though the humidity term was dropped due to multicollinearity. Some second order polynomial terms and interactions were also found to be statistically significant. The resulting model was then forced with CFSv2 forecasts of the LSEFs at leads of 1-day to over 2 months to produce several forecast products, each with different strengths:

1. Short lead forecasts:
   a. Leads of 1 and 4 days
   b. Highest POD, lowest PCA
   c. Effective for situational awareness
   d. 4 (1-day lead) or 8 (4-day lead) ensemble members

2. Weekly Outlooks:
   a. Immediate week and following week valid periods
   b. Competitive with the state of the art for operational forecast tools
   c. Capable of accounting for intraseasonal climate events such as the Madden Julian Oscillation or Kelvin waves that influence TCs but aren’t yet forecasted with skill at greater leads
   d. Ensembling is required (28 members)

3. Monthly Outlooks
   a. Includes current month, 1 and 2 month lead products
   b. The extensive ensembling yields good skill at unprecedented lead times
   c. Skill is second only to the short lead forecasts
   d. Outlooks are issued at useful lead times for government and non-government organization planners
   e. There is promise for skilled forecasts at even greater leads
   f. Most ensembling is required (120 members)
The TCFS system has been providing skilled objective forecasts for 6 years, and has been used as an input by CPC for their GTHB forecasts. It has spatial and temporal resolution, combined with long leads, that is not available in typical seasonal forecasts. Several improvements have been made along the way, as more and more is learned about TCs, formations, CFSv2, and the TCFS. A prototype TCFS for the Atlantic is scheduled to reforecast selected years of interest this winter and should be forecasting in real time, complete with reforecast verification statistics in time for the 2016 Atlantic hurricane season.

Acknowledgements:

This paper includes results from research supported by the Naval Postgraduate School Agreement No. N00244-15-2-0005 awarded by the NAVSUP Fleet Logistics Center San Diego (NAVSUP FLC San Diego). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U. S. Government.

CFSR provided by NOAA/CPC/NCEP, from their web site at http://nomads.ncdc.noaa.gov/data/cfsr/. CFSv2 data is provided at: http://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod/. JTWC best track data was provided by JTWC Tropical Cyclone Best Track Data Site at http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/shindex.html

References:


